BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

STIC-ILL

From: Portner, Ginny
Sent: Thursday, August 27, 1998 4:13 PM
To: STIC-ILL
Subject: FROM 1641

IDENTIFICATION OF CLUSTERS OF BIALLELIC POLYMORPHIC SEQUENCE-TAGGED SITES PSTSS THAT GENERATE HIGHLY INFORMATIVE AND AUTOMATABLE MARKERS FOR GENETIC LINKAGE MAPPING

%%%NICKERSON D A%%%; WHITEHURST C; BOYSEN C; CHARMLEY P; KAISER R; HOOD L
DIV. BIOL., 139-74, CALIF. INST. TECHNOL., PASADENA, CALIF. 91125.
GENOMICS 12 (2). 1992. 377-387. CODEN: GNMCE
Full Journal Title: Genomics
Language: ENG

GENOMICS 12, 377–387 (1992)

Identification of Clusters of Biallelic Polymorphic Sequence-Tagged Sites (pSTSs) That Generate Highly Informative and Automatable Markers for Genetic Linkage Mapping

DEBORAH A. NICKERSON, CHARLES WHITEHURST, CECILIE BOYSEN, PATRICK CHARMLEY, ROBERT KAISER, AND LEROY HOOD

Division of Biology, 139-74, California Institute of Technology, Pasadena, California 91125

Received August 12, 1991; revised October 15, 1991

Using a combination of denaturing gradient gel electrophoresis and direct DNA sequencing, we have found that multiple (4 to 7) biallelic sequence polymorphisms can be located within short DNA segments, 300 to 2400 bp. Here, we report on the identification of three clusters of DNA polymorphisms, one in each of the constant regions of the human T cell receptor α and β gene complexes on human chromosomes 14 and 7, respectively, and a third among the human t-RNA genes on human chromosome 14. The frequency of these polymorphisms and the extent of linkage disequilibrium between individual polymorphisms have been determined using a semiautomated DNA typing system combining DNA target amplification by the polymerase chain reaction with the analysis of internal sequence polymorphisms by a colorimetric oligonucleotide ligation assay. We have found that individual biallelic polymorphisms in each cluster are often in partial linkage disequilibrium with one another. This partial linkage disequilibrium permits the combined use of three to four markers in a cluster to generate a haplotype with high levels of heterozygosity, 71 to 88%. Therefore, clusters of physically linked biallelic polymorphisms provide an automatable and highly informative type of genetic marker for general linkage analysis as well as an attractive alternative marker system for fine-point mapping of disease-causing genes and phenotypic traits relative to their framework locations in the genome.

INTRODUCTION

The identification and detection of DNA sequence polymorphisms plays a fundamental role in understanding genome structure and function through genetic linkage mapping (Botstein *et al.*, 1980; Donis-Keller *et al.*, 1987). Many types of sequence polymorphisms are present in the genome and can be employed in genetic linkage analysis. Two major types of DNA polymorphisms stem from variations in the number of repeat units, i.e., the simple dinucleotide (CA_n/GT_n) repeats (Weber and May, 1989), and the more complex variable number tandem repeats (VNTRs) (Jeffreys *et al.*, 1986; Nakamura *et al.*, 1987). These polymorphisms frequently have multiple alleles in a population. Therefore, they have a high probability of occurring in different forms on the two copies of any given chromosome within a single individual, which makes them highly informative markers for genetic linkage mapping. Another major type of DNA polymorphism comes from discrete changes in a specific DNA sequence, i.e., single nucleotide substitutions (Botstein *et al.*, 1980). These polymorphisms are the most frequent and widely distributed type of sequence variation in the genome and are usually biallelic in the population. Individual biallelic polymorphisms are usually not as informative as polymorphic repeats. However, multiple closely linked markers can be combined into haplotypes that can be as informative as a repeat polymorphism (Donis-Keller *et al.*, 1986).

The identification and analysis of DNA polymorphisms has been greatly facilitated by the development of the polymerase chain reaction (PCR), a method that rapidly and exponentially amplifies the specific target sequences located between two oligonucleotide primers (Saiki *et al.*, 1988). PCR amplification is rapidly changing the way in which DNA analysis is performed through the development of genetic mapping strategies based on polymorphic sequence-tagged sites (pSTSs). A pSTS is any unique but short genomic sequence amplified by PCR (Olson *et al.*, 1989) that also contains an identified sequence polymorphism. Most types of DNA polymorphisms, including simple nucleotide substitutions (Feldman *et al.*, 1988), dinucleotide repeats (Weber and May, 1989), and complex VNTR repeats (Jeffreys *et al.*, 1988), can be obtained as pSTSs. However, the analysis of pSTSs, much like restriction fragment length polymorphisms (RFLPs), still relies heavily on gel electrophoresis. In addition to limiting sample throughput and the automation potential of DNA analysis, gel electrophoresis also has several disadvantages stemming from the need for appropriate internal standards and band-matching criteria to compensate for distortion problems, such as band shifting (Lander, 1991). Furthermore, the analysis of amplified DNA products by gel electrophoresis can often be complicated by the presence of artifact bands (Dracopoli and Meisler, 1990). This is

0888-7543/92 $3.00

particularly a problem with sequences containing repeats, where recombinant products can form during the amplification process (Meyerhans *et al.*, 1990). These problems make electrophoretic analysis difficult to interpret by a computer. Therefore, the development of automated and easily interpreted systems for the analysis of size variant pSTSs may be difficult. In contrast, systems for analyzing simple nucleotide substitutions, i.e., biallelic pSTSs, can be automated and easily interpreted by a computer, e.g., PCR combined with the oligonucleotide ligation assay (Nickerson *et al.*, 1990). The current major limitation of biallelic pSTSs is that individually these markers are not highly informative. In the present study, we demonstrate an interesting phenomenon related to the informativeness of multiple biallelic pSTSs. We have found that clusters of biallelic polymorphisms in partial linkage disequilibrium are located within short DNA segments, 300 to 2400 bp, and can be used to generate automatable and highly informative haplotypes with heterozygosities of 71 to 88%.

MATERIALS AND METHODS

Oligonucleotides. Amplification primers and ligation probes were assembled using standard phosphoramidite chemistry on an Applied Biosystems 380A DNA synthesizer. The sequences of all amplification primers and ligation probes are shown in Table 1. Each of the 5′ allele-specific ligation probes was biotinylated (B, Table 1) as previously described by Landegren *et al.* (1988). The adjoining 3′ reporter probes for the ligation assay were phosphorylated (P, Table 1) using "5′-phosphate-on" (Clontech) according to the manufacturers' instructions. All ligation probes were purified by reverse-phase HPLC and the phosphorylated 3′ reporter probes enzymatically labeled with digoxigenin (D, Table 1) as previously described (Nickerson *et al.*, 1990).

DNA amplification. Human genomic DNA samples were amplified from a set of 10 unrelated Caucasians, the parents from the 40 CEPH (Centre d'Etude du Polymorphisme Humaine, Paris) families, or selected CEPH families kindly provided by Dr. Richard Gatti. Genomic DNA samples (10 to 100 ng starting template) were mixed with a buffer (10 m*M* Tris–HCl, pH 8.3, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, and 0.001% gelatin) containing 40 μM of each of the four deoxynucleotide triphosphates (dATP, dGTP, dCTP, and dTTP), 0.1 μM of each of the amplification primers (Table 1), and *Taq* polymerase (25 U/ml), and amplified by 40 cycles of 20 s at 93°C, 40 s at the specified annealing temperature (Table 1), and 1 min 30 s at 72°C. Genomic DNA (10 ng) from the 80 CEPH parents was amplified for OLA analysis in 96-well microtiter plates (MJ Research, Watertown, MA). Amplification reactions were assembled as previously described (Nickerson *et al.*, 1990) using a robotic workstation (Biomek 1000, Beckman Instruments, Palo Alto, CA).

Denaturing gradient gel electrophoresis. The analysis of non-GC-clamped, amplified DNA samples by denaturing gradient gel electrophoresis (DGGE) was performed as described by Sheffield *et al.* (1989) using an electrophoretic apparatus obtained from Green Mountain Supplies (Waltham, MA). Each gel was composed of 7% acrylamide (37.5:1 acrylamide:bisacrylamide) and a gradient of denaturants (100% denaturant = 7 *M* urea and 40% (v/v) formamide) prepared in the apparatus to run either parallel or perpendicular to the electrophoretic field. Amplified DNA samples were denatured for 5 min at 100°C and cooled to room temperature for 10 min prior to gel loading to permit the formation of heteroduplexes. Amplified samples were electrophoresed for 9 h at 150 V for parallel gels and 6 h at 150 V for perpendicular gels. Following electrophoresis the gels were stained with ethidium bromide and photographed under uv transillumination. The midpoint melting temperatures (T_m) for DNA regions of interest were calculated using the Melt87 computer program generously provided by L. S. Lerman (Lerman and Silverstein, 1987) and melting maps generated by importing these data into a commercial plotting program (SigmaPlot, Jandel Scientific, Corte Madera, CA).

DNA sequencing. Amplified DNA samples were purified by electrophoresis in a 1% low melt agarose gel, and the target band excised and sequenced directly from the agarose plug using the amplification primers also as sequencing primers as described by Kretz *et al.* (1989).

Oligonucleotide ligation assay. Oligonucleotide ligation assays (OLA) for each of the described sequence polymorphisms were performed using a robotic workstation as detailed in Nickerson *et al.* (1990). Briefly, OLA employs two short (15- to 25-mers) adjacent oligonucleotide probes in the analysis of biallelic pSTSs. For each polymorphism, three oligonucleotides are synthesized, two biotinylated probes, one for each of the allelic forms of the polymorphism with the 3′ end of these probes positioned on the polymorphic nucleotide, and one adjacent 3′ reporter oligonucleotide probe common to both alleles and labeled with digoxigenin (Table 1). Analysis of an amplified DNA sample is performed using two separate ligation reactions. Each reaction contains one of the 5′ biotinylated probes, the common 3′ reporter probe, an aliquot of the amplified DNA target, and T4 DNA ligase. When the ligating probes are hybridized to a perfectly complementary target, T4 ligase can covalently join the 5′ biotinylated probe to the 3′ reporter probe. If the probes are mismatched at their target junction, T4 ligase does not form a covalent bond between the probes. For assay readout the 5′ biotinylated probe is captured on a streptavidin-coated microtiter plate and an enzyme-linked immunosorbent assay (ELISA) for digoxigenin reporter is performed. The presence (target match) or absence (target mismatch) of a colored product is measured spectrophotometrically at 490 n*M* using a microtiter plate reader. Sample genotypes are then determined by a computer program that calculates the mean absorbances from triplicate ligation reactions for each allele and then takes a ratio of these means to call a genotype (Nickerson *et al.*, 1990). Sample genotypes are transferred to a database program (dBASE III) for calculation of allele frequencies and calculation of observed haplotype frequencies. Double heterozygous individuals are excluded in the pairwise determination of observed haplotype frequencies. Linkage disequilibrium, i.e., the extent of nonrandom allelic association, between pairs of DNA polymorphisms, is calculated by using the *Q* statistic described in detail by Hedrick *et al.* (1986). The *Q* statistic is a χ^2 distributed measure of linkage disequilibrium that sums the differences between the observed haplotype frequencies and the expected haplotype frequencies (calculated by assuming random allelic association between pairs of DNA polymorphisms). The χ^2 probabilities are calculated using one degree of freedom.

RESULTS

A Highly Informative Cluster of DNA Polymorphisms Is Present in the T Cell Receptor α Chain Constant Region (Cα)

Our initial strategy for identifying DNA polymorphisms in the T cell receptor (TCR) constant region was to amplify a specific DNA segment from 10 unrelated human DNA samples and to scan these for sequence polymorphisms using parallel denaturing gradient gels. Using this strategy, we examined a 466-bp region located in the third intron of the Cα gene (Yoshikai *et al.*, 1985) and discovered a highly polymorphic and informative melting pattern consisting of four homozygous melting variants in addition to a number of heterozygous combinations of the four homozygous variants. A parallel denaturing gradient gel of these four homozygous variants and the six possible heterozygous melting variants is shown in Fig. 1A. These 10 different combinations seem

TABLE 1

Sequences of Amplification Primers and Ligation Probes

pSTS	Amplification primers	Ligation probes	
		Allele-specific probes	Reporter probes
Cα1	GAGCTAAGAGAGCCGTACTGG (55°C)[a]	B-TTAGGGACGCGGGTCTCTGC	P-GTGCATCCTAAGCTCTGAGA-D
	CTTGAAGCTGGGAGTGG	B-TTAGGGACGCGGGTCTCTGG	
Cα2	GAGCTAAGAGAGCCGTACTGG (55°C)	B-ATTTCAGCCGCCTCAGTTGA	P-ACTTCTCCCCTATGAGGTAG-D
	CTTGAAGCTGGGAGTGG	B-ATTTCAGCCGCCTCAGTTGC	
Cα3	GAGCTAAGAGAGCCGTACTGG (55°C)	B-GAACGAAGAAACTGAGGCCC	P-CACAGCTAATGAGTGGAGGA-D
	CTTGAAGCTGGGAGTGG	B-GAACGAAGAAACTGAGGCCA	
Cα4	GAGCTAAGAGAGCCGTACTGG (55°C)	B-TGTACACCACATGCCTTGTG	P-TTGTACTTCTCTCACCGTGT-D
	CTTGAAGCTGGGAGTGG	B-TGTACACCACATGCCTTGTA	
Cβ21	GGAATGGATAAGATGACTTC (60°C)	B-CTGGGATGAGGGAGACATTTTA	P-GATCTGGGGCTTCTTTGGATTTA-D
	TCCTCTTTGATGGTTCTACC	B-CTGGGATGAGGGAGACATTTTG	
Cβ22	GGAATGGATAAGATGACTTC (60°C)	B-TGGGCAGTGCCAACCCATAGGGT	P-GGATACAAAAGACAGGCAAGGAAGGG-D
	TCCTCTTTGATGGTTCTACC	B-TGGGCAGTGCCAACCCATAGGGC	
Cβ23	CATTATGGTCCTTTCCCGG (60°C)	B-ACCAGGACCAGACAGCTCTT	P-AGAGCAACCCTAGCCCCATTAC-D
	GCTTTGCTGACCCTGTGAA	B-ACCAGGACCAGACAGCTCTC	
Cβ24	CATTATGGTCCTTTCCCGG (60°C)	B-CCCAAAAGGCCACACTGGTG	P-TGCCTGGCCACAGGCTTCTA-D
	GCTTTGCTGACCCTGTGAA	B-CCCAAAAGGCCACACTGGTA	
Cβ25	GGCTCAATGTCCTTACAAAGC (60°C)	B-GACAAGAACAGAGCATGTAGGT	P-ACCTGTGCACAGGTGGCTCT-D
	AGACTCTGTTGACAAGAACAGAG	B-GACAAGAACAGAGCATGTAGGC	
Cβ26	CTGTTTCCCTGAAGATTGAG (55°C)	B-CGAAATAGGCTAAACCAATAAAAAAATT	P-GTGTGTTGGGCCTGGTTGCA-D
	AGGGCTGTCACTTCAGAGT	B-CGAAATAGGCTAAACCAATAAAAAATG	
Cβ27	CTGTTTCCCTGAAGATTGAG (55°C)	B-AGTTCTGCTCATCACTGACC	P-TATCTTCTGATTTAGGGAGCAG-D
	AGGGCTGTCACTTCAGAGT	B-AGTTCTGCTCATCACTGACT	
HT1	AAACACCAGAAAGGGTGGG (60°C)	B-GGAGAAGAAAAGGCAATATT	P-ATGTAATAACATGTAAACCCC-D
	TCAACTCAACAAGGTGTGTCT	B-GGAGAAGAAAAGGCAATATC	
HT2	AAACACCAGAAAGGGTGGG (60°C)	B-AATTGCAGCTATTAACTCCCC	P-AGGGATGTGAAGAAGCAAAGCA-D
	TCAACTCAACAAGGTGTGTCT	B-AATTGCAGCTATTAACTCCCA	
HT3	AAACACCAGAAAGGGTGGG (60°C)	B-GTTAATAGCTGCAATTTTTCAAAGAAC	P-GGTATACAGGGACAGCAAAG-D
	TCAACTCAACAAGGTGTGTCT	B-GTTAATAGCTGCAATTTTTCAAAGAAT	
HT4	CGCGATACAGACACTCTTAGG (55°C)	B-CGGCAAGAATGCGTACTTATTTGC	P-ATAGTAGAGGTAAACCACACGCC-D
	TCCTCCTACCCACAGGGATT	B-CGGCAAGAATGCGTACTTATTTGA	
HT5	CCTGAGCTATGGAGCCCTCG (58°C)	B-CTCCTAGCATTCCGGCTACT	P-GAATAGGATGTTAGCTTGAGTAAAA-D
	AATTCTTACATCAACACCGTT	B-CTCCTAGCATTCCGGCTACC	

[a] Annealing temperature for the amplification primers.

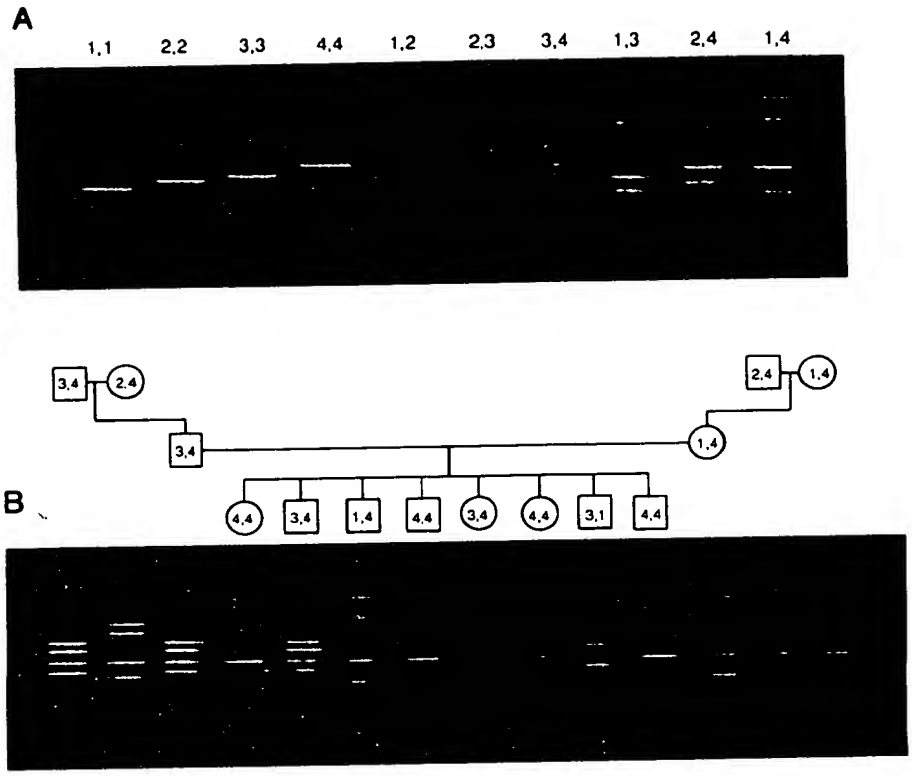


FIG. 1. DGGE analysis of an amplified Cα gene segment from intron 3. (A) An ethidium bromide-stained parallel denaturing gradient gel (0 to 80% gradient) showing the four homozygous variants (1,1; 2,2; 3,3; 4,4) and the six heterozygous variants (1,2; 2,3; 3,3; 3,4; 1,3; 2,4; 1,4) generated by the heterozygous pairing of the four homozygous forms. (B) The pedigree and DGGE genotypes of a three-generation family analyzed on a parallel denaturing gradient gel (0 to 80% gradient) using amplified DNA products obtained from each of the family members.

to represent the major forms of polymorphism for this region since we have been unable to identify any other banding patterns with the analysis of 50 additional DNA samples (CEPH parents, data not shown). The inheritance of these Cα melting variants through extended families (CEPH) has also been examined by parallel DGGE. Each variant was inherited in a Mendelian fashion (Fig. 1B).

Analysis of the amplified Cα segment by perpendicular DGGE revealed a two-domain melting structure (Fig. 2A). The presence of two melting domains was also predicted by the construction of a melting map for this Cα region and suggested the presence of a 5′ 150-bp high temperature domain and a lower temperature domain for the remaining downstream (3′) sequence (300 bp, Fig. 2B). Based on these domain properties and the ability of DGGE to detect polymorphisms located in only the lower temperature domains (Fischer and Lerman, 1983; Myers *et al.*, 1987), we predicted that the sequence change or changes responsible for the different melting variants would likely reside within the 300-bp lower temperature domain. To further evaluate this, we amplified and sequenced DNA samples from each of the homozygous melting variants (Fig. 1A). Upon comparison of these sequences we found that four biallelic nucleotide variations, Cα1 through Cα4, were responsible for generating these different melting variants. The nucleotide substitutions located at each of these polymorphic sites are shown in Fig. 3. As shown by the melting map illustrated in Fig. 2B, the most 5′ sequence polymorphism, Cα1, was located in the highest temperature domain while the remaining polymorphisms were located in the lower temperature domain. Because of its location in the higher temperature domain, we suspected that the Cα1 polymorphism was not responsible for any of the mobility shifts detected by DGGE. This was later confirmed by the observation that DNA samples heterozygous at Cα1, but homozygous at Cα2, Cα3, and Cα4, form a single homozygous band on a parallel denaturing gradient gel (data not shown). In contrast to Cα1, the other sequence polymorphisms (located in the lower temperature domain) did appear to affect the relative mobility pattern of the Cα segment when analyzed by parallel DGGE. More importantly, the relative migration patterns for each of these melting variants in the denaturing gel corresponded directly to the number of GC versus AT basepairs at the three downstream polymorphic sites. Note that variants having more GC basepairs at these polymorphic sites migrated further into the gel (compare Figs. 1A and 3). Therefore, we surmise that each replacement of an AT with a GC basepair served to further stabilize the lower temperature domain, resulting in a slightly higher melting temperature and, as a consequence, further migration into the gel.

To rapidly obtain estimates of the allele frequencies for each of these polymorphisms in a population, we utilized a nonisotopic semiautomated method, PCR/OLA (Nickerson *et al.*, 1990), to detect each of the polymor-

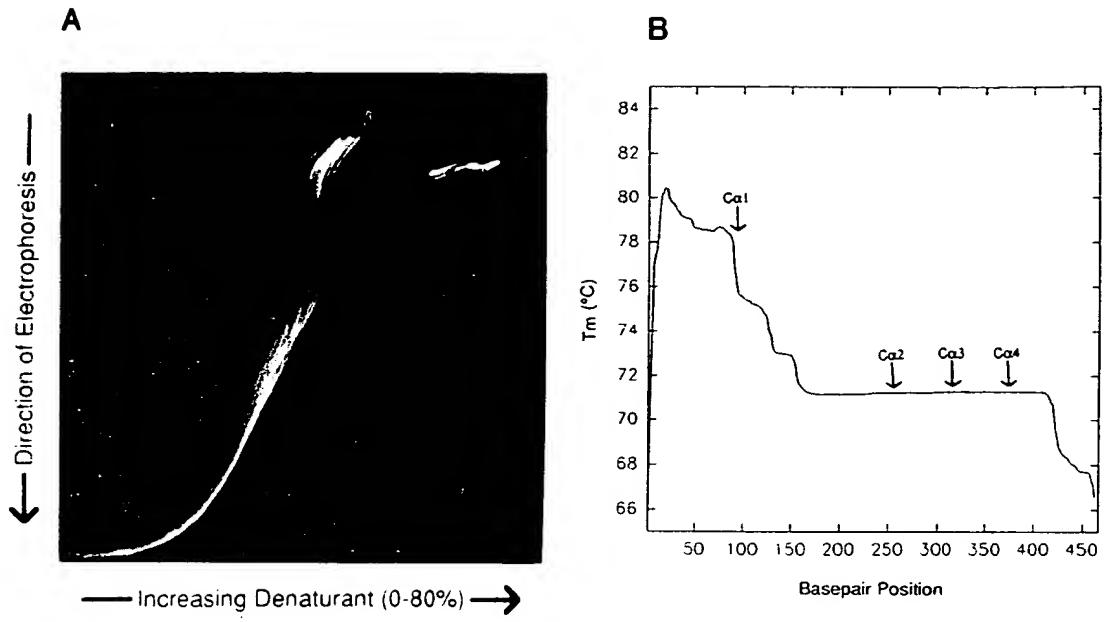


FIG. 2. Analysis of the Cα gene segment. (A) An ethidium bromide-stained perpendicular denaturing gradient gel using amplified DNA products obtained from a pooled DNA template (equivalent amounts of DNA from 20 unrelated individuals) showing both the domain structure for this region and the large number of homozygous and heteroduplex alleles that can be detected in this highly informative region. (B) Calculated melting map of the amplified Cα fragment showing the relative location of the four nucleotide substitutions in this gene segment.

phic alleles at these four sites. An example of these results with DNA samples from homozygous and heterozygous individuals is shown in Fig. 3. Clearly, the PCR/OLA procedure is a highly effective approach for discriminating each of the allelic forms at these four sites. Analysis of a population of 80 individuals (the CEPH parents) by PCR/OLA revealed that all four Cα polymorphisms had genotype distributions indicating Hardy–Weinberg equilibrium. A pairwise analysis of linkage disequilibrium (nonrandom association) between these polymorphisms showed a moderate level of disequilibrium ($P = 8.3 \times 10^{-3}$ to 9.8×10^{-6}) between most polymorphisms with the exception of the Cα1 and Cα3 pair which revealed a much higher level of linkage disequilibrium in this population ($P = 5.9 \times 10^{-17}$, Table 2). Although each polymorphism revealed a moderate degree of heterozygosity among the 80 CEPH parents, 25 to 41% (Table 2), the degree of heterozygosity increased markedly when these polymorphisms were combined into a haplotype. The combined heterozygosity of

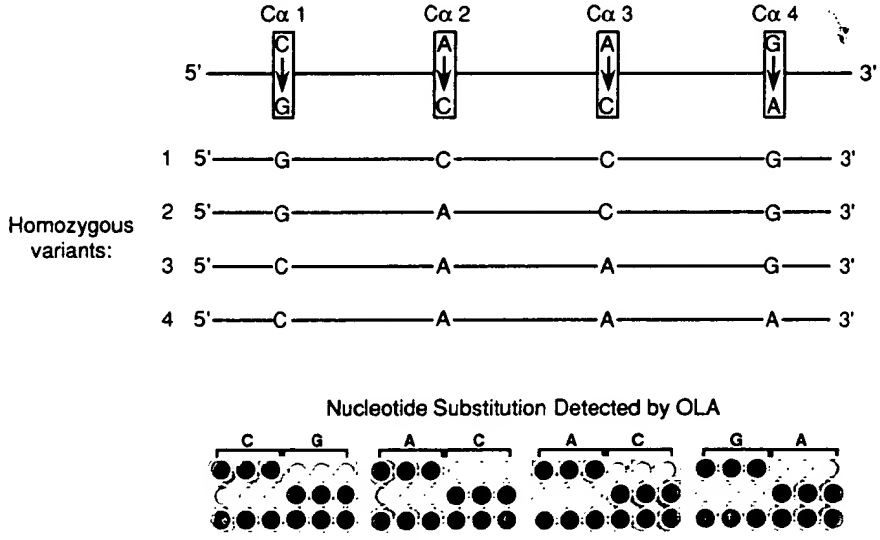


FIG. 3. Sequence and OLA analysis of the Cα polymorphisms. (**Top**) A schematic diagram indicating the nucleotide variations identified by sequencing amplified DNA samples from each of the four homozygous DGGE variants. (**Bottom**) ELISA-based OLA analysis of the four Cα sequence polymorphisms using amplified DNA samples from homozygous and heterozygous individuals. OLAs for the each allele were performed in triplicate. Microtiter wells containing the digoxigenin reporter form a colored product and indicate complete complementarity between the ligating probes and amplified DNA target. The absence of a colored product in the microtiter wells indicates a mismatch between the ligating probes and the amplified DNA target.

TABLE 2

Allele Frequency, Heterozygosity, and Linkage Disequilibrium in the Cα pSTS Cluster

	Cα1	Cα2	Cα3	Cα4
Nucleotide	3480	3645	3702	3754
Allele frequency	C: 68% G: 32%	A: 86% C: 14%	A: 32% C: 68%	G:69% A: 31%
Observed heterozygosity	41%	25%	40%	39%
Linkage Disequilibrium				
Cα1	—	$P = 3.7 \times 10^{-5}$ [a]	$P = 5.9 \times 10^{-17}$	$P = 1.2 \times 10^{-5}$
Cα2	17.8, $n = 130$ [b]	—	$P = 2.6 \times 10^{-5}$	$P = 8.3 \times 10^{-3}$
Cα3	70.0, $n = 96$	17.7, $n = 132$	—	$P = 9.8 \times 10^{-6}$
Cα4	19.1, $n = 136$	7.0, $n = 148$	19.6, $n = 136$	—

[a] P values for the χ^2 distribution of the Q statistic.
[b] Values for Q and the number of chromosomes examined.

the four polymorphisms was 71% by PCR/OLA and was identical to the heterozygosity obtained by combining only three of the polymorphisms, Cα1/Cα2/Cα4 or Cα2/Cα3/Cα4.

A Cluster of DNA Polymorphisms Located in the TCR β Constant Region 2 (Cβ2)

To further explore the extent of DNA sequence polymorphisms in the constant regions of the TCR, we used a strategy similar to that described for Cα to examine a 483-bp segment from the TCR Cβ2 region (104 bp from the 5′ noncoding region and 379 bp from exon 1; Toyonaga *et al.*, 1985). A biallelic polymorphism within this Cβ2 region was detected by analysis of amplified DNA from 10 unrelated individuals on a parallel denaturing gradient gel. The inheritance of this Cβ2 polymorphism, shown in Fig. 4, followed a Mendelian pattern. Sequence analysis of amplified DNA samples from the two homozygous melting variants indicated that a single nucleotide substitution in the lower melting domain of the Cβ2 sequence was responsible for generating the different melting variants (Cβ23, a C to T base change, Fig. 5). However, like Cα, the complete sequence of this segment revealed the presence of another DNA polymorphism in the highest melting domain (Cβ24, Fig. 5). This polymorphism resides within the coding region of Cβ2 (exon 1) but does not lead to any amino acid substitutions. Although DGGE has proven effective in detecting DNA polymorphisms within the lower melting domains of these amplified fragments, we have found that direct DNA sequencing offers increased sensitivity in the detection of DNA polymorphisms. Furthermore, we have found that direct DNA sequencing of amplified DNA targets even with multiple individuals (e.g., 6 unrelated individuals) is more rapid than DGGE analysis since DNA sequencing: (i) can be performed under a standard set of conditions that offers considerable time-savings, (ii) will unambiguously determine whether a common DNA polymorphism (at least 1 in 6 individuals) is present across the entire scanned region (high or low melting domains), and (iii) provides the information necessary for high throughput typing and automated data analysis by approaches like OLA. In contrast, several types of gels, i.e., a perpendicular and/or parallel gels (0 to 80% and 30 to 80% gradients), must be explored to properly

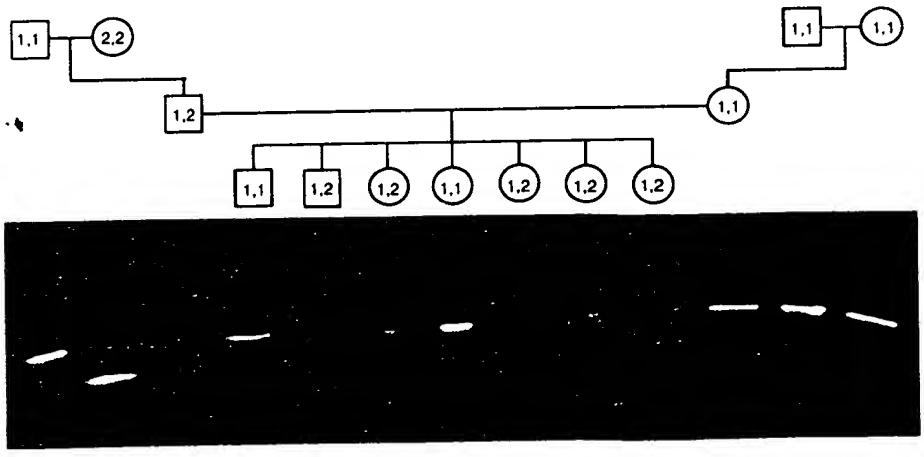
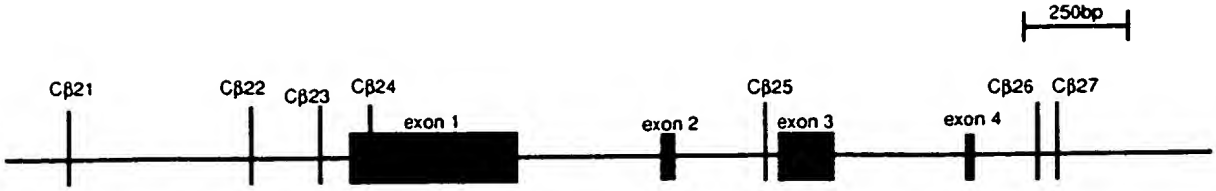


FIG. 4. DGGE analysis of the Cβ2 gene segment. The pedigree and DGGE genotypes of a three generation family analyzed by a parallel denaturing gradient gel (30 to 80% gradient) with amplified Cβ2 products obtained from each of the family members.



Nucleotide:	167	581	745	871	1,820	2,467	2,525
Allele Frequency:	T: 45% C: 55%	C: 48% T: 52%	C: 62% T: 38%	G: 47% A: 53%	A: 52% G: 48%	T: 58% G: 42%	C: 51% T: 49%
Observed Heterozygosity:	55%	54%	40%	49%	51%	45%	53%
Linkage Disequilibrium:							
Cβ21	——	$p=1.3 \times 10^{-6}$ [a]	$p=8.3 \times 10^{-5}$	$p=5.2 \times 10^{-4}$	$p=1.1 \times 10^{-2}$	$p=5.3 \times 10^{-3}$	$p=1.8 \times 10^{-3}$
Cβ22	23.4, n=98 [b]	——	$p=7.0 \times 10^{-11}$	$p=1.1 \times 10^{-5}$	$p=5.8 \times 10^{-4}$	$p=6.8 \times 10^{-4}$	$p=9.8 \times 10^{-5}$
Cβ23	15.5, n=116	42.5, n=102	——	$p=6.6 \times 10^{-7}$	$p=1.1 \times 10^{-4}$	$p=1.3 \times 10^{-4}$	$p=3.8 \times 10^{-6}$
Cβ24	12.0, n=108	19.3, n=104	24.7, n=112	——	$p=8.3 \times 10^{-7}$	$p=6.0 \times 10^{-7}$	$p=1.9 \times 10^{-7}$
Cβ25	6.4, n=116	11.8, n=112	14.9, n=118	24.3, n=102	——	$p=1.4 \times 10^{-11}$	$p=1.3 \times 10^{-14}$
Cβ26	7.8, n=118	11.6, n=112	14.7, n=124	24.9, n=106	45.7, n=96	——	$p=4.0 \times 10^{-12}$
Cβ27	9.7, n=108	15.2, n=106	21.4, n=114	27.1, n=100	59.3, n=78	48.1, n=94	——

a. p values for chi-square distribution of Q.
b. Values for Q, and the number of chromosomes examined.

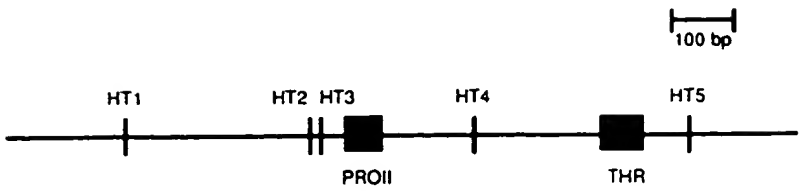
FIG. 5. The location, frequency, and linkage disequilibrium among the biallelic DNA polymorphisms found in the Cβ2 gene segment of the human TCR.

scan even the lower temperature domain of a specific DNA sequence by DGGE. Furthermore, even when a DNA polymorphism has been detected by DGGE, sufficient information is not obtained to permit high throughput typing or analysis. Therefore, on the basis of the ease, sensitivity, and speed of direct DNA sequencing, we modified our approach to polymorphism scanning. Using a sequence-based approach, we examined an additional 1900 bp of noncoding sequence from Cβ2. Five additional single nucleotide variations were identified in Cβ2 using a sequence-based approach (Cβ21, Cβ22, Cβ25, Cβ26, and Cβ27 designated according to their location in Cβ2 5′ to 3′, Fig. 5). The majority of these additional sequence polymorphisms (four of five) were the result of transitional base substitutions (Cβ21, Cβ22, Cβ25, and Cβ27). Moreover, two of these polymorphisms provide the underlying sequence variations responsible for two previously reported RFLPs in Cβ2, the *Bgl*II polymorphism (Cβ21; Berliner *et al.*, 1985), and the *Kpn*I polymorphism (Cβ25; Perl *et al.*, 1989).

Using a combination of PCR and OLA (Table 1) we have confirmed that each of these seven Cβ2 polymorphisms followed a Mendelian inheritance pattern. Figure 5 shows the location, allele frequencies, and heterozygosities for these polymorphisms when analyzed by PCR/OLA for 80 DNA samples (the CEPH parents). Six of the seven polymorphisms had allele frequencies approximating 50:50 while the remaining polymorphism, Cβ23, revealed an obvious major and minor allele pattern (approximately 1:2). All polymorphisms had genotype distributions that were not significantly different from those calculated assuming Hardy–Weinberg equilibrium. Furthermore, we find that these seven polymorphism are in varying degrees of linkage disequilibrium with one another (Fig. 5). On the basis of the moderate levels of disequilibrium present in this region, a highly informative haplotype can be generated using only four or five of these markers (heterozygosity of 88% for the Cβ21/Cβ22/Cβ24/Cβ25 combination and 89% for the Cβ21/Cβ22/Cβ24/Cβ25/Cβ26 combination). As would be predicted, pairs of polymorphisms exhibiting the highest levels of disequilibrium, e.g., Cβ22/Cβ23, Cβ25/Cβ26, Cβ25/Cβ27, and Cβ26/Cβ27, contributed the least to the overall informativeness of a haplotype from this region.

A Cluster of Sequence Polymorphisms Is Found in the Human ProII and Thr t-RNA Genes

Our findings of multiple physically linked but highly informative clusters of biallelic polymorphisms in the TCR constant regions (Cα and Cβ2) led us to question whether this finding was a common phenomenon in the genome or specific to the TCR gene complexes. To assess this, we selected a known sequence of moderate length (approximately 1300 bp) surrounding the *Pro*II and Thr t-RNA genes located on human chromosome 14 (Chang *et al.*, 1986). Three sequence-tagged sites (STSs) were generated from the noncoding sequences surrounding these t-RNA genes (each approximately 300 to 450 bp in length). Using amplified DNA samples from six unrelated individuals and our direct DNA sequencing approach, we detected five single nucleotide substitutions within these STSs, three transitions and two



NUCLEOTIDE:	194	475	506	763	1,110
ALLELE FREQUENCY:	T: 18% C: 82%	G: 79% T: 21%	C: 78% T: 22%	C: 41% A: 59%	T: 22% C:78%
OBSERVED HETEROZYGOSITY:	28%	38%	41%	42%	36%
LINKAGE DISEQUILIBRIUM:					
HT1	——	$p=6.7 \times 10^{-3}$[a]	$p=1.0 \times 10^{-2}$	$p=5.6 \times 10^{-4}$	$p=3.3 \times 10^{-2}$
HT2	7.4, n=146[b]	——	$p=1.3 \times 10^{-11}$	$p=2.2 \times 10^{-5}$	$p=4.0 \times 10^{-9}$
HT3	6.5, n=148	45.7, n=98	——	$p=2.1 \times 10^{-4}$	$p=1.9 \times 10^{-9}$
HT4	11.9, n=136	18.0, n=110	13.8, n=112	——	$p=2.1 \times 10^{-5}$
HT5	4.5, n=146	34.6, n=104	36.1, n=104	18.1, n=108	——

a. p values for the chi-square distribution of the Q statistic.
b. Values for Q, and the number of chromosomes examined.

FIG. 6. The location, frequency, and linkage disequilibrium for the biallelic DNA polymorphisms identified in DNA sequences surrounding the human pro II and thr t-RNA genes.

transversions as shown in Fig. 6. These polymorphisms have been designated HT1 through HT5 according to their 5′ to 3′ location within the sequence (Fig. 6) and were all inherited in a Mendelian fashion. Each of these polymorphisms showed a moderate level of heterozygosity (24 to 42%) and had genotype distributions indicating Hardy–Weinberg equilibrium when tested on a panel of 80 DNA samples (the CEPH parents) by PCR/OLA. Moreover, like the Cα and Cβ2 polymorphisms, several of the HT polymorphisms also showed a moderate level of linkage disequilibrium with one another (Fig. 6). Therefore, three polymorphisms could be combined to generate a highly informative haplotype for this region, HT1/HT3/HT4 (combined heterozygosity, 72%); this seems to indicate that clusters of multiple linked biallelic polymorphisms may be a common phenomenon in the genome.

DISCUSSION

Single Nucleotide Substitutions Are Predominantly Transitions

We have detected the presence of clusters of sequence polymorphisms in the human genome using a scanning approach employing DGGE and/or direct DNA sequencing. Altogether, 16 single nucleotide substitutions were uncovered, 4 in the TCR Cα region, 7 in the TCR Cβ2 region, and 5 among the sequences surrounding two human t-RNA genes. All of these polymorphic substitutions were inherited in a Mendelian fashion and yielded allele distributions similar to those expected assuming Hardy–Weinberg equilibrium. Furthermore, these substitutions also followed the general trends for the types of nucleotide substitutions that occur within the human genome (Vogel and Kopun, 1977). Specifically, transitional base changes appear more frequently than nucleotide transversions. Additionally, transitional base changes involving C and T are clearly favored over substitutions involving G and A. Among the nucleotide substitutions found surrounding the human t-RNA genes, and in the TCR Cα and Cβ2 regions, 10 of the 16 substitutions (62%) were transitional base changes, and substitutions involving C and T were the most frequent type of transitional change (70%).

DNA Sequence Polymorphisms Are Frequent

Many estimates on the frequency of polymorphism in the human genome have been reported. These estimates vary depending on the number of individuals examined and the technique used to scan the DNA segment, but they generally range from 1 in every 200 bp to 1 in every 1000 bp (Cooper *et al.*, 1985; Miyamoto *et al.*, 1988). The overall distribution of DNA polymorphisms in the three regions we examined was 16 nucleotide variations in a total of 3755 bp of DNA, or 1 in every 235 bp. Considering the wide range of variation that may exist in the distribution of sequence polymorphisms within the human genome, our observation of a polymorphism distribution of 1 in every 235 bp among predominantly (90%) noncoding sequences probably represents a fairly neutral distribution of base substitution (Kimura, 1983).

It is noteworthy that only 5 of the 16 substitutions in these three regions would have been identified as RFLPs, one of the most common scanning approaches for identifying DNA polymorphisms. Even denaturing gradient gels (in the absence of a GC-clamp) detected only a moderate number of DNA polymorphisms in the regions scanned (four of six polymorphisms among the

Cα and Cβ segments scanned). This level of sensitivity using several variables for the analysis, perpendicular and parallel gels, i.e., 0 to 80% and 30 to 80% gels, is consistent with other regions we have scanned by direct DNA sequencing and DGGE (random chromosome 14 STSs and mouse TCR variable gene segments, C. Boysen, M. Toda, and D. A. Nickerson, unpublished observations). The most accurate method for identifying DNA polymorphisms is of course direct DNA sequence analysis. Until recently, however, this type of scanning was tedious and mainly restricted to well-defined regions of the genome (Miyamoto *et al.*, 1988). The speed and efficiency of PCR amplification combined with new approaches for direct sequence analysis of amplified DNA products (Kretz *et al.*, 1989; Rosenthal and Jones, 1990) provides a rapid and highly accurate system for scanning for DNA sequence polymorphisms as demonstrated here and previously by Yandell and Dryja (1989). Furthermore, sequence-based approaches to polymorphism scanning can be automated to generate high throughput systems for analyzing amplified DNA samples from multiple individuals (Wahlberg *et al.*, 1990; Wilson *et al.*, 1990).

DNA Polymorphisms Are Frequently in Partial Linkage Disequilibrium

We have found that DNA polymorphisms within an individual cluster, Cα, Cβ and t-RNA, were in partial linkage disequilibrium with one another. This is similar to findings for several other regions of the human genome, i.e., the α-anti-trypsin and insulin genes (Cox *et al.*, 1985; Chakravarti *et al.*, 1986). In the present study, DNA polymorphisms (Cα3 and Cα4) separated by only 52 bp were found to be in partial linkage disequilibrium. Furthermore, it is apparent that some of the polymorphisms within these clusters remain in high levels of linkage disequilibrium (Cα1 and Cα3) while others found between (Cα2 and Cα4) or on either side of these polymorphisms exhibit only partial disequilibrium. Many factors could play a role in generating partial linkage disequilibrium between physically linked DNA polymorphisms including random genetic drift as well as the rate of recombination between these polymorphisms over time in a population (Ohta, 1982; Thomson and Klitz, 1987). However, among DNA polymorphisms separated by such small physical distances (<2000 bp), it is possible that the levels of linkage disequilibrium could reflect the evolutionary history of these polymorphisms over time in the population. For example, DNA polymorphisms arising through a series of sequential mutational events could give rise to multiple haplotypes in the population (Fig. 7). In contrast, pairs of DNA polymorphisms arising more recently, or in tandem, i.e., closely timed mutations, might be at higher levels of disequilibrium with one another in the population (Fig. 7, haplotype III). The presence of multiple haplotypes and partial disequilibrium in clusters of DNA polymorphisms is clearly a phenomenon that can be exploited for

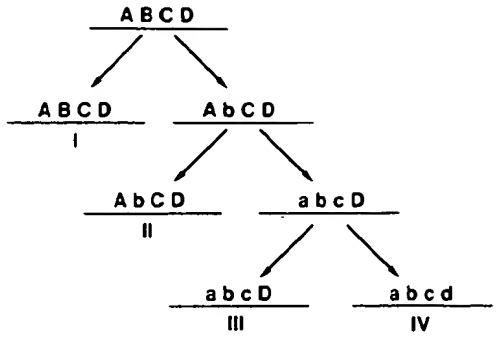


FIG. 7. A schematic diagram for the formation of multiple DNA haplotypes generated in the population through a series of sequential single base mutations (haplotypes II and IV) or through tandem (pairs of closely timed or recently generated) mutations (haplotype III) as suggested from the analysis of the TCR Cα haplotypes.

the development of alternative forms of highly informative markers, i.e., clusters of pSTSs. Furthermore, partial linkage disequilibrium between DNA polymorphisms in short DNA segments may well be a common finding in the genome since we have also detected partial disequilibrium among polymorphisms in DNA segments obtained by random cloning methods (data not shown).

PCR/OLA Provides a High Throughput Genetic Mapping Procedure

The development of high throughput systems for the identification and typing of sequence polymorphisms will greatly increase the speed and efficiency of genetic linkage mapping as well as the identification of the genes encoding traits of interest, such as disease-causing genes. In this regard and as our data show, pSTSs containing simple biallelic nucleotide substitutions offer significant advantages in the rapid development of linkage maps. First, single nucleotide substitutions are the most common and widely distributed type of polymorphism in the genome. Therefore, it is likely that any random STS obtained from the genome would have a high probability of containing one or more single nucleotide substitutions depending on the size of the STS. Second, once identified, biallelic pSTSs can be typed using a rapid and semiautomatable system, PCR/OLA. Finally and more importantly, the outcome from the analysis of biallelic pSTSs by PCR/OLA is easy to interpret unequivocally as a positive or negative by a computer, a feature not easily achieved with size variant markers analyzed by gel electrophoresis, i.e., repeat polymorphisms.

Clustered Polymorphisms in pSTSs Are Highly Informative

The most useful genetic markers are those that are sufficiently informative to discriminate all four parental chromosomes. More informative genetic markers reduce the number of meioses required to map a specific trait and help to map genetic diseases with limited affected pedigrees. In this regard, the use of clusters of simple biallelic polymorphisms offers an alternative highly in-

formative marker system for genetic linkage mapping. For example, once a disease-causing gene is mapped to a defined framework interval (±5 cM), the identification of clusters of highly informative biallelic markers would provide an easily attained marker for localizing a disease-causing gene(s) mapped to this interval. Additionally, the concept of identifying clusters of pSTSs can also be extended to large fragments of human DNA. For example, once a random biallelic pSTS with a single base substitution has been identified, the PCR primers as well as the OLA probes (both allele-specific probes combined with the reporter probe) can be used to rapidly screen an ordered array of YAC or cosmid clones to identify a larger physical fragment of human DNA containing this pSTS (P. Kwok and M. Olson, unpublished observation). This will provide a physical anchor for each pSTS as well as a source of linked DNA sequence to develop additional STSs for chromosome walking in a defined framework interval or to develop additional pSTSs markers to generate a highly informative marker at a defined location in the genome, i.e., in framework locations surrounding a disease-causing gene or phenotypic trait. Therefore, the development of clusters of biallelic pSTSs can serve as a means of connecting the physical and genetic maps of the genome in addition to providing an alternative, automatable, and highly informative marker system for genetic linkage mapping.

ACKNOWLEDGMENTS

We thank Mr. Stephen Lappin for his technical assistance; Dr. Suzanna Horvath for her assistance in oligonucleotide synthesis; Dr. Ben Koop for his helpful discussions; and Drs. L. Clawson, T. Hunkapiller, L. Martin, and T. Yager for their review and comments on this manuscript. This work was supported by National Institutes of Health Grants HG 00084 and HG 00464, the Whittier Foundation, and National Science Foundation Grant DIR 8809710.

REFERENCES

Berliner, N., Duby, A. D., Morton, C. C., Seidman, J. G., and Leder, P. (1985). Detection of a frequent restriction fragment length polymorphism in the human T cell antigen receptor beta chain locus. *J. Clin. Invest.* **76**: 1283–1285.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* **32**: 314–331.

Chakravarti, A., Elbein, S. C., and Permutt, M. G. (1986). Evidence of increased recombination near the human insulin gene: Implication for disease association studies. *Proc. Natl. Acad. Sci. USA* **83**: 1045–1049.

Chang, Y-N., Pirtle, I. L., and Pirtle, R. M. (1986). Nucleotide sequence and transcription of a human tRNA gene cluster with four genes. *Gene* **48**: 165–174.

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S., and Schmidtke, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69**: 201–205.

Cox, D. W., Woo, S. L. C., and Mansfield, T. (1985). DNA restriction fragments associated with α_1-antitrypsin indicate a single origin for deficiency allele PIZ. *Nature* **316**: 79–81.

Donis-Keller, H., Barker, D., Knowlton, R., Schumm, J. W., Braman, J. C., and Green, P. (1986). Highly polymorphic RFLP probes as diagnostic tools. *Cold Spring Harbor Symp. Quant. Biol.* **51**: 317–322.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephen, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lander, E. S., Botstein, D., Akots, G., Rediker, K. S., Gravius, T., Brown, V. A., Rising, M. B., Parker, C., Powers, J. A., Watt, D. E., Kauffman, E. R., Brecker, A., Phipps, P., Muller-Kahle, H., Fulton, T. R., Ng, S., Schummn, J. W., Braman, J. C., Knowlton, R. G., Barker, D. F., Crooks, S. M., Lincoln, S. E., Daly, M. J., and Abrahamson, J. (1987). A genetic linkage map of the human genome. *Cell* **51**: 319–337.

Dracopoli, N. C., and Meisler, M. H. (1990). Mapping the human amylase gene cluster on the proximal short arm of chromosome 1 using a highly informative $(CA)_n$ repeat. *Genomics* **7**: 97–102.

Feldman, G. L., Williamson, R., Beaudet, A. L., and O'Brien, W. E. (1988). Prenatal diagnosis of cystic fibrosis by DNA amplification of KM-19 polymorphism. *Lancet* **ii**: 102.

Fischer, S. G., and Lerman, L. S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory. *Proc. Natl. Acad. Sci. USA* **80**: 1579–1583.

Hedrick, P. W., Thomson, G., and Klitz, W. (1986). Evolutionary genetics: HLA as an exemplary system. *In* "Evolutionary Processes and Theory" (S. Karlin and E. Nevo, Eds.), pp. 583–606, Academic Press, New York.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J., and Ponder, B. A. J. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Amer. J. Hum. Genet.* **39**: 11–24.

Jeffreys, A. J., Wilson, V., Neuman, R., and Keyte, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16**: 10,953–10,971.

Kimura, M. (1983). "The Neutral Theory of Molecular Evolution," Cambridge Univ. Press, Cambridge.

Kretz, K. A., Geoffrey, S. C., and O'Brien, S. O. (1989). Direct sequencing from low-melt agarose with Sequenase. *Nucleic Acid Res.* **17**: 5864.

Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988). A ligase-mediated gene detection technique. *Science* **241**: 1077–1080.

Lander, E. S. (1991). Research on DNA typing catching up with courtroom application. *Am. J. Hum. Genet.* **48**: 819–823.

Lerman, L. S., and Silverstein, I. (1987). Computational simulation of DNA melting and its applications to denaturing gradient gel electrophoresis. *In* "Methods in Enzymology" (R. Wu, Ed.), Vol. 155, pp. 482–502, Academic Press, San Diego.

Meyerhans, A., Vartanian, J-P., and Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Res.* **18**: 1687–1691.

Miyamoto, M. M., Koop, B. F., Slightom, J. L., Goodman, M., and Tennant, M. R. (1988). Molecular systematics of higher primates: Genealogical relations and classification. *Proc. Natl. Acad. Sci. USA* **85**: 7627–7631.

Myers, R. M., Maniatis, T., and Lerman, L. S. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. *In* "Methods In Enzymology" (R. Wu, Ed.), Vol. 155, pp. 501–527, Academic Press, San Diego.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, H., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987). Variable number tandem repeat (VNTR) markers for human gene mapping. *Science* **235**: 1616–1622.

Nickerson, D. A., Kaiser, R., Lappin, S., Stewart, J., Hood, L., and Landegren, U. (1990). Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proc. Natl. Acad. Sci. USA* **87**: 8923–8927.

Ohta, T. (1982). Linkage disequilibrium due to random genetic drift in

finite subdivided populations. *Proc. Natl. Acad. Sci. USA* **79:** 1940–1944.

Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* **245:** 1434–1435.

Perl, A., Divincenzo, J. P., Gergely, P., Condemi, J. J., and Abraham, G. N. (1989). Detection and mapping of polymorphic *Kpn*I alleles in the human T-cell receptor constant beta-2 locus. *Immunology* **67:** 135–138.

Rosenthal, A., and Jones, D. S. C. (1990). Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. *Nucleic Acids Res.* **18:** 3095–3096.

Saiki, R. K., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., and Erlich, R. B. (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.

Sheffield, V. C., Cox, D. R., Lerman, L. S., and Myers, R. M. (1989). Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-based changes. *Proc. Natl. Acad. Sci. USA* **86:** 232–236.

Thomson, G., and Klitz, W. (1987). Disequilibrium pattern analysis. I. Theory. *Genetics* **116:** 623–632.

Toyonaga, B., Yoshikai, Y., Vadasz, V., Chin, B., and Mak, T. W. (1985). Organization and sequences of the diversity, joining, and constant region genes of the human T-cell receptor β chain. *Proc. Natl. Acad. Sci. USA* **82:** 8624–8626.

Vogel, F., and Kopun, M. (1977). Higher frequencies of transitions among point mutations. *J. Mol. Evol.* **9:** 159–180.

Wahlberg, J., Lundeberg, J., Hultman, T., and Uhlen, M. (1990). General colorimetric method for DNA diagnostics allowing direct solid-phase genomic sequencing of the positive samples. *Proc. Natl. Acad. Sci. USA* **87:** 6569–6573.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 388–396.

Wilson, R. K., Chen, C., and Hood, L. (1990). Optimization of asymmetric polymerase chain reaction for rapid fluorescent DNA sequencing. *Biotechniques* **8:** 184–189.

Yandell, D. W., and Dryja, T. P. (1989). Detection of DNA sequence polymorphisms by enzymatic amplification and direct genomic sequencing. *Am. J. Hum. Genet.* **45:** 547–555.

Yoshikai, Y., Clark, S. P., Taylor, S., Sohn, V., Wilson, B. I., Minden, M. D., and Mak, T. W. (1985). Organization and sequences of the variable, joining and constant region genes of the human T-cell receptor α-chain. *Nature* **316:** 837–840.